

## Research Article

# 4.2 Million and Counting... The Animal Toll for REACH Systemic Toxicity Studies

Jean Knight<sup>1</sup>, Thomas Hartung<sup>2,3</sup> and Costanza Rovida<sup>2</sup>

<sup>1</sup>Consultant, Half Moon Bay, CA, USA; <sup>2</sup>Center for Alternatives to Animal Testing Europe (CAAT-Europe), University of Konstanz, Konstanz, Germany;

<sup>3</sup>Center for Alternatives to Animal Testing (CAAT), Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, USA

### Abstract

The EU's chemicals regulation, REACH, requires that most chemicals in the EU be evaluated for human health and ecosystem risks, with a mandate to minimize use of animal tests for these evaluations. The REACH process has been ongoing since about 2008, but a calculation of the resulting animal use is not publicly available. For this reason, we have undertaken a count of animals used for REACH. With EU legislators set to consider REACH revisions that could expand animal testing, we are releasing results for test categories counted to date: reproductive toxicity tests, developmental toxicity tests, and repeated-dose toxicity tests for human health. The total animal count as of December 2022 for these categories is about 2.9 million. Additional tests involving about 1.3 million animals are currently required by a final proposal authorization or compliance check but not yet completed. The total, 4.2 million, for just these three test categories exceeds the original European Commission forecast of 2.6 million for all REACH tests. The difference is primarily because the European Commission estimate excluded offspring, which are most of the animals used for REACH. Other reasons for the difference are extra animals included in tests to ensure sufficient survive to meet the minimum test requirement; dose range-finding tests; extra test animal groups, e.g., for recovery analysis; and a high rejection rate of read-across studies. Given higher than forecast animal use, the upcoming debate on proposed REACH revisions is an opportunity to refocus on reducing animal numbers in keeping with the REACH mandate.

## 1 Introduction

The Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Regulation (EC 1907/2006) was enacted to manage human health and ecosystem risks of chemical substances manufactured or imported in quantities of 1 ton/year or more. The regulation also established the European Chemicals Agency (ECHA) to administer the regulation and ensure compliance.

REACH specifies human health and ecosystem effects for which substances must be evaluated. A defined effect, such as short-term repeated-dose toxicity, is called an *endpoint*, and the required endpoints depend on the annual tonnage of the substance. These requirements are given in REACH Annexes VII (for 1–10 tons/year), VIII (for 10–100 tons/year), IX (for 100–1,000 tons/year), and X (for > 1,000 tons/year), with the requirements increasing with tonnage range.

Companies registering existing substances were required to submit dossiers with the chemical information, including the toxicity evaluations, to ECHA by 2010 (Annex X substances),

2013 (Annex IX substances), or 2018 (Annex VII and VIII substances). The 2010 deadline also applied to substances classified “carcinogenic, mutagenic or toxic for reproduction category I” and substances classified “persistent, bioaccumulative and toxic or very persistent and very bioaccumulative” in the tonnage band  $\geq 100$  tons/year. New toxicity tests needed to meet the first deadline, 2010, typically had to be performed without delay. From personal experience, we are confident that few tests were conducted for REACH before 2009 due to some confusion regarding REACH implementation and the 2008 financial crisis, which induced companies to request a postponement of the deadline (which was not granted). We therefore consider 2009 as the starting year for counting animal tests for REACH. ECHA, too, writes: “2009 is taken as a significant point in time, as it defined the studies that generally should be conducted and motivated by the REACH requirements...” (ECHA, 2020).

REACH Articles 13 and 25 mandate animal testing only as a last resort; therefore, many substances have toxicity data from available existing toxicity tests and from *read-across* studies,

Received March 20, 2023; Accepted June 13, 2023;  
© The Authors, 2023.

ALTEX 40(3), 389–407. doi:10.14573/altex.2303201

Correspondence: Costanza Rovida, PhD  
CAAT-Europe  
University of Konstanz, Box 600  
78464 Konstanz, Germany  
costanza.rovida@chimici.it

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



which use existing tests from similar substances to estimate data for the registered substance, according to the provisions specified for read-across in REACH. If no reliable existing tests are available for the substance itself or for a read-across study, the substance generally undergoes a new experimental test. This may be an alternative (non-animal) experiment or an animal experiment, depending on the endpoint and substance properties.

Our goal is to count the animals used for all REACH endpoints. To date, we have completed the count for endpoints in the following categories: repeated-dose toxicity, reproductive toxicity, and developmental toxicity. Because EU legislators will soon be debating REACH revisions, we are releasing the results for these categories early to help inform the debate. The results are for the following six endpoints:

- Short-term repeated-dose toxicity (Annex VIII)
- Sub-chronic repeated-dose toxicity (Annexes IX, X)
- Reproductive/developmental toxicity screening (Annexes VIII, IX)
- Multi-generation reproductive toxicity (Annex X, potentially Annex IX if deemed necessary based on substance characteristics and toxicity data)
- Developmental toxicity in a first species, usually rat (Annexes IX, X)
- Developmental toxicity in a second species, usually rabbit (Annex X, potentially Annex IX if deemed necessary based on substance characteristics and toxicity data)

These endpoints have no non-animal methods accepted by regulators, so if existing or read-across data are not available or are rejected by ECHA, any new experiment is an animal experiment.

New animal experiments typically involve more than just the main test. Besides the main test, they may include animals for one or more of the following purposes:

- Maximum tolerated dose or palatability tests, conducted prior to the main test to find the highest workable dose.
- Dose range-finding tests, conducted prior to the main test to determine suitable low, medium, and high doses for the main test.
- Extra groups such as recovery groups (to assess reversibility, persistence, or delayed occurrence of toxic effects) or groups for evaluating a specific effect, such as liver toxicity.
- Spare animals, purchased in case replacements are needed for any reason.

Multi-generation tests also involve the offspring of the parent animals. Generally, all animals are sacrificed at the end of the study, even unused spare animals.

The substance dossiers include much of this animal data, entered into specific fields of the ECHA database. ECHA makes non-confidential portions of the dossiers available to the public through its online database<sup>1</sup>. This database and its known field format are what allowed us to systematically extract the animal count data.

After the European Commission (EC, 2001) proposed the REACH strategy in 2001, studies estimated new testing needs and consequent animal use (Pedersen et al., 2003; van der Jagt

et al., 2004; Hofer et al., 2004; Rovida and Hartung, 2009). In a report for the European Commission, Pedersen et al. (2003) estimated the number of tests that would be conducted for each proposed endpoint. Van der Jagt et al. (2004) then used the Pedersen et al. (2003) test numbers to derive a European Commission estimate for expected animal use, which in the average case was a total of 2.6 million animals.

At the time, no standard method existed for counting animals in experimental tests, but a common method, and the one used by van der Jagt et al. (2004), was to count only animals used in the main test, excluding offspring in multi-generation tests. Hofer et al. (2004) and Rovida and Hartung (2009) also counted only animals in the main tests but included offspring for some or all multi-generation tests. This is a major reason that their total animal estimates of 7.5 million (Hofer et al., 2004) and 54.4 million (Rovida and Hartung, 2009) were much higher than the van der Jagt et al. (2004) estimate. The Rovida and Hartung (2009) estimate was also based both on the number of pre-registered substances and on the expected market increase in the number of substances, which indicated many more substances would be registered than initially estimated by Pedersen et al. (2003). Ultimately, registrations have been lower than the Pedersen et al. (2003) estimate.

Since REACH took effect, ECHA has periodically estimated the number of new experimental tests that could relate to REACH (ECHA, 2017a, 2020). In the ECHA estimates, a test was counted if the test was a guideline study consistent with REACH, was performed on the registered substance, and had a study year of 2009 or later. The algorithms used by ECHA screened for duplicate tests, which are common within and across dossiers, but the algorithms could not detect some duplicates (ECHA, 2020). The algorithms also did not screen out tests for purposes other than REACH, such as for publications or foreign regulations. ECHA noted that these limitations meant their numbers were likely overestimated (ECHA, 2017a, 2020). Despite these limitations, these numbers have been the best estimates of REACH-related tests to date. Taylor (2018) used these numbers to calculate REACH animal use from 2009-2016, estimating 2.2 million animals for completed and pending tests for that period.

Like ECHA, we use an algorithm to identify new experimental tests in the REACH dossiers, and the screening parameters are similar but include more. The test must be a guideline study consistent with REACH, performed on the registered substance or a form of the substance, with a study year of 2009 or later. Further, tests obtained from publications are excluded, and free-text fields commonly used to note a non-REACH purpose are flagged if the text suggests such a purpose. The algorithm also includes more dossier fields for identifying duplicate tests.

We obtained the animal data directly from the dossiers in the ECHA public database, which provides the most accurate count. It also gives insight into actual practice, which is useful for estimates going forward. For example, reproductive toxicity test guidelines typically recommend using “*sufficient number of mat-*

<sup>1</sup> <https://www.echa.europa.eu/information-on-chemicals/registered-substances>

**Tab. 1: Endpoints included in the animal count**

This paper presents the animal count for the repeated-dose, reproductive, and developmental toxicity endpoints shown in this table. The table shows the specific endpoint required by REACH and the test methods used for evaluating the endpoint if a test is needed.

REACH section <sup>a</sup>	Standard requirement	OECD test guideline <sup>b</sup>
<b>8.6</b>	<b>Repeated-dose toxicity endpoints</b>	
8.6.1	Short-term (28-day) repeated-dose toxicity	407 (oral) 410 (dermal) 412 (inhalation)
8.6.2	Sub-chronic (90-day) repeated-dose toxicity	408 (oral) 411 (dermal) 413 (inhalation)
<b>8.7</b>	<b>Reproductive and developmental toxicity endpoints</b>	
8.7.1	Screening for reproductive/developmental toxicity	421 or 422
8.7.2	Pre-natal developmental toxicity, first species (rat)	414
8.7.2	Pre-natal developmental toxicity, second species (rabbit)	414
8.7.3	Two-generation reproductive toxicity <sup>c</sup>	416
8.7.3	Extended one-generation reproductive toxicity study (EOGRTS) <sup>c</sup>	443
N/A	One-generation reproduction toxicity study (Not a standard requirement but sometimes used for a reproductive toxicity endpoint. OECD deleted this test method in 2019.)	415

<sup>a</sup> REACH Annexes VIII through X (see Section 8 in each of those annexes): <https://eur-lex.europa.eu/legal-content/en/TXT/HTML/?uri=CELEX:02006R1907-20220501#tocId249>; <sup>b</sup> OECD test guidelines are internationally agreed upon test methods for assessing chemical safety: <https://www.oecd.org/chemicalsafety/testing/oecdguidelinesfortheassessmentofchemicals.htm>; <sup>c</sup> Commission Regulation (EU) 2015/282 requested the extended one-generation reproductive toxicity study instead of the two-generation toxicity study. This took effect in March 2015. After that, a two-generation study was accepted only if it was initiated before 13 March 2015 (ECHA, 2017b).

ing pairs to yield at least 20 pregnant females.” The dossier data reveal what “sufficient” translates to in real numbers.

In 2010, EU Directive 2010/63/EU adopted a new counting standard that includes all animals that “are used or intended to be used in procedures, or bred specifically so that their organs or tissues may be used for scientific purposes” (Article 1). Our count follows the 2010 Directive, so it includes animals in the main test – including offspring – plus extra groups, spare animals, and dose range-finding animals; and for rabbit studies, it also includes animals for palatability/maximum tolerated dose tests.

A new, important revision of REACH is currently under discussion in the EU. The draft version is expected by the end of 2023. Two proposals would expand animal testing:

- Extending the requirement for a chemical safety assessment (CSA) to Annex VII substances, i.e., those registered in the tonnage band 1–10 tons/year, which currently are exempt from this requirement. Currently, the CSA for human health is based on animal studies. Many Annex VII substances do not have an animal study, so a new animal study, typically a combined repeated-dose reproductive toxicity screening study, would be requested for them.
- Registration of polymers, a large group of substances that are currently exempted.

A recent amendment of REACH (Regulation EU 2022/477) already expands the request for new animal tests, for example, by

asking for an animal test for genotoxicity in case of a positive result from an *in vitro* test and by expanding the conditions to request an animal study on fish for the assessment of the bioconcentration factor. This regulation also reduces the possibilities to waive the extended one-generation reproductive toxicity test (EOGRTS) and the developmental toxicity test performed on the second species in Annex X as discussed in Rovida et al. (2023) in this issue.

As EU legislators debate further expanding requests for new animal testing, understanding the current state of REACH animal testing is important. This paper presents the current state of animal use for repeated-dose toxicity, reproductive toxicity, and developmental toxicity endpoints, which account for most animal use in REACH. It provides a direct estimate of animal use plus related statistics that may be helpful in the debate. Our animal count will continue with the remaining endpoints for REACH, and we welcome suggestions for improving the accuracy of the count.

## 2 Method

Table 1 shows the specific endpoints and the associated animal tests that are included in this animal count. The method involved identifying substances with such tests conducted or pending for REACH and extracting the dossier data for the substances to a database in a format that allowed the animals to be counted.



## 2.1 Identification of substances with animal tests

The public ECHA database has limited search options. To overcome this difficulty, substances with possible REACH animal tests were identified by searching the eChemPortal database<sup>2</sup>. The eChemPortal database contains chemical information entered for government chemical programs, including REACH. Its search feature allows selection of studies by specified parameters, such as endpoint and study year. At the time of our search, REACH data in eChemPortal was current as of 14 November 2022, the date ECHA selected their REACH data for upload to the eChemPortal database.

Using the eChemPortal Property Search feature, we searched the portal for ECHA REACH experimental studies for “Toxicity to reproduction”, “Development toxicity/teratogenicity” and “Repeated-dose toxicity” dated 2009 or later and with the highest reliability: “1 (reliable without restriction)”. Our assumption is that most studies newly done for REACH are guideline studies with high reliability. Manual review of a sample of 100 studies of reliability “2 (reliable with restriction)” indicated about 5% of reliability 2 studies could be relevant to REACH. The other reliability 2 studies in the sample were literature, read-across studies, dose range-finding tests, ECHA inquiry results/summaries, or clearly not for REACH. Note that excluding reliability 2 studies from the eChemPortal search did not affect our ability to get dose range-finding tests, because those were obtained during the dossier reviews for the main studies (Section 2.2). Requiring a reliability of 1 was a good screening mechanism, and the tradeoff of a potential 5% undercount of reliability 2 studies was considered acceptable.

The search results identified over 4,000 unique REACH substance dossiers meeting the criteria. Note that this does not mean that all dossiers contained new animal tests for REACH. The eChemPortal search results do not identify whether the test is done on the registered substance or a different, similar substance as part of a read-across approach; or whether the test was originally done for a purpose other than REACH. That information was determined during the dossier review process described in Section 2.2.

## 2.2 Review and selection of tests for the animal count

Each substance dossier identified in the eChemPortal search was reviewed to assess whether tests for the specified endpoint were performed on the substance for REACH. Tests were selected for the animal count if they met the following criteria:

- Report date of 2010 or later, with study year of 2009 or later. If no report date was given, the criterion was a study year of 2009 or later.
- Reliability of 1 for main tests.
- Test material was the registered substance or a form (e.g., hydrate) of the substance used instead of the substance. If the test material or registered substance was not identified and no clar-

ifying information could be found in the dossier, the test was handled as follows:

- *Registered substance known, test material unknown*: Test counted if it was the only test included for the endpoint or if it used OECD TG 422 (a test not widely used outside REACH); otherwise, test not counted.
- *Registered substance unknown, test material known*: Test counted if all other endpoints in the dossier used the same test material and if the test material did not also appear as a test material in other dossiers (indicating possible read-across); otherwise, test not counted.
- *Registered substance and test material both unknown*: Test not counted.

Usually, when registered substance or test material was unknown, the dossier contained clarifying information and the preceding decision process was unnecessary.

- List of test methods included an OECD or EU test guideline appropriate for REACH. If the only test method listed was a method specific to a non-EU country (e.g., an EPA method), it was assumed to be done for a purpose other than REACH.
- Species was appropriate for REACH.
- Reference type was not a publication.
- No indication of read-across or non-REACH purpose.

An initial batch of dossiers was manually reviewed in the public ECHA database<sup>3</sup>. Based on this initial review, we identified key dossier fields for evaluating whether a test met the selection criteria. We then created a Python<sup>4</sup> program to save those fields for all remaining dossiers in an Excel spreadsheet and to flag potential issues (e.g., test material = registered substance?), allowing efficient review of tests. Tests unclear from the Python data (e.g., tests where the registered substance or test material was not identified in the saved fields) were manually reviewed directly in the dossiers available in the public ECHA database.

Although eChemPortal identified individual tests in a dossier, all tests for the endpoint of interest in the dossier were reviewed by the algorithm to identify tests that may have been added since the last eChemPortal update and to find dose range-finding tests for the main studies.

The saved data included fields where registrants might note if a test was for a purpose other than REACH. To further screen for non-REACH tests, we checked an ECHA report that listed tests prior to 2014 with a non-REACH purpose (ECHA, 2014). Likely, our count includes some non-REACH tests, because registrants are not required to note this in the dossiers.

Duplicate tests are prevalent within and across dossiers and excluding them is important to avoid overestimating animal use. Duplicate tests typically occurred when the same OECD TG 422 test was entered for repeated-dose toxicity and reproductive toxicity endpoints or when the test for one registered substance was used for one or more other substances but was not identified in

<sup>2</sup> <https://echemportal.org/echemportal/property-search> (accessed and search results downloaded 6 April 2022 and 24 December 2022 (for repeated-dose toxicity), 5 May 2022 and 17 December 2022 (for reproductive toxicity), and 14 July 2022 and 17 December 2022 (for developmental toxicity)).

<sup>3</sup> <https://echa.europa.eu/information-on-chemicals/registered-substances/> (accessed April-May 2022).

<sup>4</sup> Python is an open-source programming language well-suited to automating routine tasks (<https://www.python.org/>).



those other dossiers as read-across or an analogue study. The Python data let us identify duplicate studies by comparing test materials and often unique details, such as exact study periods or in-life dates, across studies. Some studies provided little identifying information, and duplicates among those may still be included.

Dossier pages for tests meeting the criteria were saved from the ECHA online database<sup>5</sup>. The file content was the html page source code in text format, to allow extraction of the animal data by other Python programs. Dose range-finding tests for the main tests were also saved if found in the dossier.

### 2.3 Count of animals

For tests that met the selection criteria, a second Python program scanned the saved files, extracted dossier fields relevant to the animal count, and saved the data to an Excel spreadsheet for calculation of animal numbers, available in the supplementary file<sup>6</sup>. The following fields were saved (field names given as they appear in the public ECHA database):

- Substance identifier fields: Substance name, EC number, CAS number
- Experiment fields: Test method, Limit test, Route of administration, Species, Sex, No. of animals per sex per dose, Control animals (negative control), and Positive control. Fields containing dose information were also extracted to count number of doses.

#### 2.3.1 Main test animals

Most dossiers had animal data for the main tests. For the few tests with only partial animal data, missing parameters (e.g., number of animals/sex/dose) were estimated as the average value for the known tests.

#### 2.3.2 Offspring

Reproductive toxicity and developmental toxicity tests include parent animals (F0 generation) and their offspring (F1 generation and sometimes a third, F2, generation). The number of parent animals was well documented in dossiers, but the number of offspring was not. Few dossiers reported the total number of pups born. If number of pups was reported, the number usually included only pups used to create the F1 and F2 treatment groups, and excluded the pups who were culled.

If total number of pups, including culled pups, was given, that number was used. Otherwise, pups were estimated based on average litter size and estimated number of litters.

- *Average litter size*: We used dossier data on litter size rather than published litter size estimates for which we were unsure of the basis. Dossier data provided representative litter sizes across control, low, medium, and high doses, and also reflected tests, especially those done for compliance checks, that were on priority substances more likely to have toxic effects

potentially reducing litter size. Litter sizes were found for 38 reproductive and developmental toxicity tests using rats and 18 developmental toxicity tests using rabbits. Since these tests involve about 40 to 80 or more litters, depending on the test, the tests provided a large sample set: over 1,900 rat litters and over 1,100 rabbit litters. Average litter size was 9.0 pups for rabbits and 11.6 pups for rats, which were rounded to the nearest whole numbers: 9 and 12.

- *Estimated number of litters*: Estimated from the initial number of paired females for a test, which was reported for most tests, and average *fertility index*, which is the number of females paired for mating who become pregnant. Fertility index was found for 142 dose groups for rats and 73 dose groups for rabbits. Average fertility index was 94% for rabbits and 92% for rats. Based on limited dossier data on loss of litters during gestation, this number was adjusted to 90% for both species, a somewhat arbitrary adjustment but included to be conservative. In other words, 90% of paired females were estimated to deliver litters.

Number of pups was then estimated as *estimated number of litters* multiplied by *average litter size*. For example, if a test started with 40 paired female rats (10/dose group, 4 dose groups), the number of pups was estimated as (0.90 x 40) litters x 12 pups/litter = 432 pups.

#### 2.3.3 Dose range-finding animals

Dose selection information was available in most dossiers in one of the following ways:

- A dose range-finding test was included in the dossier as a separate study entry.
- A description of the dose selection method (e.g., use of historical data or dose range-finding test) was included in the “Details on study design” field for the main test.

*Dose range-finding tests in dossiers*: If a dose range-finding test was included as a separate study, range-finding animals were directly counted from the dossier information. Many range-finding tests involved offspring, and they were estimated in the same manner described in Section 2.3.2, except the adjusted fertility index was 80%, reflecting a death rate in dose range-finding tests that is often higher than in main tests. This likely occurs because data for choosing suitable doses are often less available for these preliminary tests. Note that tests initially conducted as a main test and later used for dose range-finding for a different test were counted only for their initial purpose as a main test, and the range-finding animals were zeroed out for the later test.

*Dose range-finding information in “Details on study design” field*: Dose range-finding tests identified in this field typically did not provide enough animal data to count the animals. For example, the number of doses might be given, but not the number of animals per dose. For these dose range-finding tests, number

<sup>5</sup> <https://echa.europa.eu/information-on-chemicals/registered-substances/> (accessed May-December 2022).

<sup>6</sup> doi:10.14573/altex.2303201s



of animals was estimated as the average animal count for known dose range-finding tests for that test method and species. If dose selection was based on other methods (e.g., historical data), range-finding animals for the main test were set to zero.

**Dose selection method unknown:** For main tests with no dose selection information, we assumed the unknown methods included the same percentage of dose range-finding tests as the known methods. For example, if 100 main tests had an unknown dose selection method, and 65% of main tests with known dose selection method used a dose range-finding test, then we assumed 65 of the 100 main tests with unknown method also used a dose range-finding test.

**Other preliminary tests:** Some studies included a maximum tolerated dose test *in lieu* of or in addition to a dose range-finding test. Our animal count does not include maximum tolerated dose tests except in the case of OECD TG 414 studies conducted on rabbits. A maximum tolerated dose test is almost always conducted in this case and typically is 12 animals. This number was added to the count for all OECD TG 414 rabbit studies.

Range-finding tests sometimes were preceded by a smaller animal test (e.g., using 3 to 12 animals) to find suitable doses for the range-finding test. In effect, these small tests are dose range-finding tests for the main dose range-finding test. Our animal count does not include these early dosing tests.

### 2.3.4 Spare animals

Spare animals are kept in case needed to replace animals for any reason. Dossiers rarely reported the number of spare animals. For the few that did, the number varied widely, from 2 to a number larger than the number of actual test animals. For those not reported, which was most, we estimated spare animals as 10% of the number of main test animals (parent animals for multi-generation tests), because the reported numbers were typically 10%–25%. We welcome suggestions for improving the accuracy of this parameter.

## 2.4 Identification of pending animal tests

Some substances have animal tests scheduled or under way, but the tests were not yet present in the dossier at the time of our analysis. These tests primarily result from ECHA decisions related to the following procedures:

- **Testing proposal evaluation:** New tests to fulfill the requirements of Annexes IX and X require formal authorization from ECHA. The idea is to prevent new tests if existing data are available. The testing proposal is included in the registration dossier, and most are approved 180 days to 2 years after the proposal is submitted (statutory deadlines dictate the evaluation period). In many cases, the test takes three years to complete; therefore, many tests approved from testing proposals submitted in 2018 or later were not yet available in the public ECHA database at the time of our analysis.

- **Compliance check:** Each year, ECHA selects some dossiers for in-depth assessment, including adequacy of the toxicological data. Compliance checks focus on priority endpoints, which include those in our analysis: repeated-dose toxicity, developmental toxicity, and reproductive toxicity. Compliance checks often result in a requirement for more animal tests. As with test proposals, there is typically a long lag time from ECHA's decision to when the test is completed.

- **Substance evaluation:** ECHA and the Member State authorities select dossiers that enter the Community Rolling Action Plan (CoRAP) for further evaluation. All data are strictly reviewed, with the review focused on the concern that a specific use may have. New test requests typically are not standard information requirements. The program has resulted in only 9 requests for sub-chronic, developmental, and reproductive toxicity tests since 2018<sup>7</sup> and is not considered in our analysis.

To identify pending tests, we used the ECHA Dossier Evaluation Status list<sup>8</sup>, which gives the status of all testing proposal evaluations and compliance checks and provides links to each decision, including test details. We manually reviewed the decisions from 2018–2022, the likely period when tests could still be pending, and identified decisions requesting new animal tests on a substance. We excluded tests that were annulled or withdrawn by ECHA on appeal and tests related to registrations that are no longer valid (e.g., due to Brexit). Read-across decisions were also excluded unless the test was required on the registered substance as part of a category read-across decision.

ECHA's test requests were then compared with the completed tests we had saved from the dossiers to identify requested tests that are not yet present in dossiers. These would be either pending or no longer necessary, for example, if an adaptation such as read-across was provided instead of the test and the adaptation was accepted by ECHA.

A test was counted as pending if:

- the dossier noted that the test was under way, or
  - the dossier had an adaptation but ECHA had status as “Follow up”, indicating ECHA did not yet consider the dossier compliant, or
  - the dossier now had the new test, but it was not present in the dossier when we extracted our data and so was not yet counted.
- If the dossier had an adaptation and ECHA had status as “Concluded” (indicating ECHA considered the dossier compliant), the adaptation was considered to be accepted by ECHA and the test to be no longer necessary.

Note that use of adaptations can be significant. For example, for the 429 2018–2019 compliance check decisions requesting new sub-chronic, developmental, or reproductive toxicity tests and for which the deadline for submitting the test had passed, registrants provided a new test for 67% and an adaptation (usually read-across) for 26%; the other 7% were unclear. These percentages are not necessarily representative of other years, be-

<sup>7</sup> <https://echa.europa.eu/further-information-requests-2018>, <https://echa.europa.eu/further-information-requests-2019>, <https://echa.europa.eu/further-information-requests-2020>, <https://echa.europa.eu/further-information-requests-2021>, <https://echa.europa.eu/further-information-requests-2022>

<sup>8</sup> <https://www.echa.europa.eu/information-on-chemicals/dossier-evaluation-status>; list downloaded 29 January 2023.

**Tab. 2: Total animal count for tests found in dossier review**

Total for main tests includes extra groups and spare animals. For multi-generation tests, it includes the animals from all generations. Note that the totals here cannot be used to calculate average number of animals per main test or dose range-finding (DRF) test because the totals here include test variations for both main tests and DRF tests, and each variation has a different average. For average number of animals per main test with each variation considered, see Table 5. For DRF test averages, see Table 3.

OECD test guideline	No. of tests <sup>b</sup>	Total animals used in main tests	No. of main tests with DRF test <sup>d</sup>	Total animals used in DRF tests	Total animals
<b>Reproductive toxicity</b>					
416	44	112,521	19	9,574	122,095
415	13	15,764	6	3,082	18,846
443	79	104,101	44	22,916	127,017
443 w F2 <sup>a</sup>	53	117,986	33	16,264	134,250
421	543	307,208	256 <sup>e</sup>	12,079	319,287
422	1,243	710,824	1,162 <sup>e</sup>	46,765	757,589
<b>Subtotal</b>	<b>1,975</b>	<b>1,368,404</b>	<b>1,520</b>	<b>110,680</b>	<b>1,479,084</b>
<b>Developmental toxicity</b>					
414-rat	782	907,815	550	164,433	1,072,248
414-rabbit	197	173,299	194	44,066 <sup>f</sup>	217,365
414-mouse	2	2,410	0	0	2,410
<b>Subtotal</b>	<b>981</b>	<b>1,083,524</b>	<b>744</b>	<b>208,499</b>	<b>1,292,023</b>
<b>Sub-chronic repeated-dose toxicity</b>					
408	586 <sup>c</sup>	58,084	441	15,161	73,245
411	9 <sup>c</sup>	748	7	238	986
413	89 <sup>c</sup>	11,212	72	2,612	13,824
<b>Subtotal</b>	<b>684</b>	<b>70,044</b>	<b>520</b>	<b>18,011</b>	<b>88,055</b>
<b>Short-term repeated-dose toxicity</b>					
407	422 <sup>c</sup>	22,314	357	12,058	34,372
410	19 <sup>c</sup>	1,218	14	437	1,655
412	77 <sup>c</sup>	5,439	61	2,179	7,618
<b>Subtotal</b>	<b>518</b>	<b>28,971</b>	<b>432</b>	<b>14,674</b>	<b>43,645</b>
<b>Total</b>	<b>4,158</b>	<b>2,550,943</b>	<b>3,216</b>	<b>351,864</b>	<b>2,902,807</b>

<sup>a</sup> OECD TG 443 has variations that depend on the endpoints requiring investigation. One variation requires the mating of the F1 generation to produce the F2 generation. <sup>b</sup> Tests with report date of 2010 or later, with study period of 2009 or later. <sup>c</sup> Total tests include tests without recovery groups and tests with recovery groups. See Section 3.1.3 for details. <sup>d</sup> Includes number of known DRF tests plus number of DRF tests estimated for main tests that did not report dose selection information. See Section 2.3.3 for the estimation method. <sup>e</sup> Total DRF tests include tests with offspring and tests without. See Section 3.1.2 for details. <sup>f</sup> For rabbit studies, includes DRF animals plus animals used for an initial maximum tolerated dose test, which typically uses 12 animals/test. We assumed all rabbit studies include such a test. For the 197 rabbit tests, the total for animals for the maximum tolerated dose tests is 2,364 animals.

DRF = dose range-finding

cause ECHA's priorities for compliance checks are not the same every year. The 2018-2019 percentages are given only to illustrate that adaptations must be considered when estimating tests based on compliance check and test proposal decisions.

Further analysis and excellent discussion on testing proposals can be found in Taylor et al. (2014).

### 3 Results

A total of 4,158 tests for human health systemic toxicity tests were found in 2,699 unique dossiers. Sufficient dossier information to count animals was present for over 99% of repeated-dose toxicity tests. For reproductive toxicity and developmental tox-

**Tab. 3: Average number of animals for dose range-finding tests**

Dose range-finding (DRF) tests with complete animal data were used to calculate average animals per DRF test. The DRF tests were either individual studies with their own record in the dossier or from "Dose selection rationale" information in the "Details on study design field" (Section 2.3.3).

OECD test guideline for main study	No. of DRF tests with full animal data	Total animals in DRF tests with full animal data	Avg. animals/ DRF test
<b>Reproductive toxicity</b>			
443/416/415 <sup>a</sup>	38	19,324	508
421/422 (DRF w/ F1) <sup>b</sup>	39	9,193	236
421/422 (DRF w/o F1) <sup>b</sup>	281	8,975	32
<b>Developmental toxicity</b>			
414—rat	131	39,152	299
414—rabbit	71	15,257	215
<b>Repeated-dose toxicity</b>			
407/408	103	3,521	34
410/411 <sup>c</sup>	3	63	Insufficient data; use 407/408 avg.
412/413	36	1,299	36
<b>Total</b>		<b>96,784</b>	

<sup>a</sup> OECD TG 416 and 415 tests had only four and two, respectively, DRF tests with animal data. They were combined with the OECD TG 443 DRF tests to obtain a more meaningful average for them. DRF tests for all three methods used a similar range-finding method, typically a modified OECD TG 421 or 422. <sup>b</sup> DRF tests for OECD TG 421 and 422 were combined because they used the same two methods. One method used mated females and had offspring (DRF w/ F1 in table); the other method used unmated animals and so had no offspring (DRF w/o F1 in table). These are shown separately due to the large difference in animal use between the methods. <sup>c</sup> OECD TG 410 and 411 (dermal repeated-dose) studies had only three DRF tests with animal data. We assume dermal DRF tests will have animal use similar to DRF tests for other repeated-dose studies and so have assigned the oral average to be the dermal average, too.  
DRF = dose range-finding

icity, parent animals could be counted for over 99% and 94% of tests, respectively, and offspring were estimated from other reported parameters. The results are presented here.

### 3.1 Total animal count

The total count of animals for all tests that met our criteria is about 2.9 million, which includes animals used for the main tests, extra groups added for further analysis, dose range-finding tests, and spare animals. For multi-generation tests, the total includes the animals from all generations. Totals by endpoint are summarized in Table 2. The following sections provide more details underlying the total counts in Table 2.

#### 3.1.1 Main test animals

Main test animals total about 450,000 adult animals and about 2.1 million pups. Pups account for about 72% of the total 2.9 million animals. This is an expected result, because most tests counted here are reproductive toxicity or developmental toxicity studies involving one or two generations of offspring.

#### 3.1.2 Dose range-finding animals

Dose range-finding tests accounted for about 12% of total animal use. No guidelines exist for dose range-finding tests, and they

varied widely in their method, but the following tests were more common:

- Reproductive toxicity tests: An OECD TG 421 or a modified OECD TG 421 with fewer animals.
- Reproductive toxicity screening tests and repeated-dose toxicity tests: A 14-day repeated-dose test using 3 to 5 animals/sex/dose and 2 to 3 treatment groups. About 5% of the screening tests used dose range-finding tests with mated females and offspring.
- Developmental toxicity tests: A test using 4 to 10 pregnant females and 3 treatment groups.

Most main studies (65%) included the dose selection method, so it was possible to estimate the frequency of use of dose range-finding tests. Of the main studies that identified a dose selection method, 77% used a dose range-finding test, either alone or together with existing data.

Study types with the highest use of dose range-finding tests were OECD TG 422 (94%), OECD TG 414 (70% for rat studies and 98% for rabbit studies), OECD TG 443 (63%), and the repeated-dose tests (73%-85%). Study types with the lowest use were OECD TG 416 (43%) and OECD TG 421 (47%).

Dose range-finding tests with complete animal data were used to calculate average dose range-finding animals for each OECD test guideline. The test data and resulting averages are shown in



**Tab. 4: Estimated dose range-finding tests**

While Table 3 shows animal totals for dose range-finding (DRF) tests with known animal data, this table shows totals for DRF tests that had to be estimated because the animal data were unknown. This table includes (1) known DRF tests that had no or only partial animal data and (2) estimated DRF tests for main studies with no dose selection information. For both, the animal estimate is obtained by multiplying the number of tests by the average animals/DRF test for the corresponding main study type (Tab. 3). The table also includes the estimated animals for maximum tolerated dose (MTD) tests for rabbit studies, a preliminary test normally conducted for these studies. The total in this table (255,080) plus the total in Table 3 (96,784) equals the total DRF animals in Table 2 (351,864).

OECD test guideline	No. of DRF tests with incomplete animal data	Estimated no. of DRF tests for studies with no dose selection information <sup>a</sup>	Total estimated no. of DRF animals <sup>b</sup>
<b>Reproductive toxicity</b>			
415	1	3	2,032
416	9	7	8,128
443	24	20	22,352
421	93 (4 w F1) <sup>c</sup>	113 (7 w F1) <sup>c</sup>	8,836
422	519 (1 w F1) <sup>c</sup>	373 (15 w F1) <sup>c</sup>	31,840 <sup>d</sup>
<b>Developmental toxicity</b>			
414-rat	209	210	125,281
414-rabbit	57	66	26,445
		197 (MTD) <sup>e</sup>	2,364
<b>Repeated-dose toxicity</b>			
407	154	166	10,880
408	205	170	12,818 <sup>f</sup>
410	4	7	374
411	1	6	238
412	20	27	1,692
413	24	26	1,800
<b>Total for DRF tests with estimated animal numbers</b>			<b>255,080</b>

<sup>a</sup> Based on the percentage of DRF tests for main studies with known dose selection method. See Section 2.3.3 for the estimation method.

<sup>b</sup> Estimated by multiplying number of DRF tests by average animals/DRF test for that OECD study type (Tab. 3). <sup>c</sup> For OECD TG 421 and 422 studies, most DRF tests involved unmated animals with no offspring, using an average of 32 animals; but some involved mated females with offspring (F1), using an average of 236 animals. <sup>d</sup> Number of main studies with a DRF test was 519, but one study had two DRF tests, so total includes animals for one additional DRF test (32 animals). <sup>e</sup> Assumes each OECD TG 414-rabbit study had an MTD test with 12 animals/test.

<sup>f</sup> Number of main studies with a DRF test was 205, but one study had three DRF tests, so total includes animals for two additional DRF tests (68 animals).

DRF = dose range-finding; MTD = maximum tolerated dose

Table 3. These averages were then used to estimate number of animals for dose range-finding tests with incomplete animal data. They also were used in estimating number of dose range-finding animals for studies with no dose selection information (Section 2.3.3). Table 4 shows these estimated animal numbers.

The 2004 estimate of REACH animal use assumed no dose range-finding tests would be needed, i.e., that existing data would be accepted for the dose selection method (van der Jagt, 2004). As 77% of studies reporting this data included a dose range-finding test, this is one significant difference to the original REACH estimates. For studies not using a dose range-finding test, the selection method was usually historical data (e.g., previous tests),

but some used the limit dose as the high dose and then used intervals recommended in the guideline for low and medium doses. Dossiers reported only the chosen dose selection method, not the reason for the choice, so the reason for the high use of dose range-finding tests, including when historical data were available and also used, is unknown.

### 3.1.3 Extra animal groups

Studies may include extra animal groups beyond those used for the main test. The most common groups in tests analyzed here are recovery groups for repeated-dose studies and groups to evaluate lung burden and bronchoalveolar lavage fluid in repeated-dose

**Tab. 5: Average number of animals per main test**

This shows the average number of animals per test with and without a dose range-finding (DRF) test. Note that you cannot obtain the total animal count by multiplying the last column by the total number of tests for each test method, because not all main studies used a DRF test. For total animal count, see Table 2. Test methods with the same average no. of DRF test animals used the same type of DRF test (Tab. 3).

OECD test guideline	No. of tests	Avg. no. of animals/ main test	Avg. no. of animals/ DRF test	Total avg. animals/ test if DRF used
415	13	1,213	508	1,721
416	44	2,590 <sup>b</sup>	508	3,098
443	79	1,318	508	1,826
443 with F2 ext. <sup>a</sup>	53	2,226	508	2,734
421	543	566	236/32 <sup>c</sup>	802/598 <sup>c</sup>
422	1,243	572	236/32 <sup>c</sup>	808/604 <sup>c</sup>
414-rat	782	1,161	299	1,460
414-rabbit	197	880	215	1,095
407, without recovery groups	297	44	34	78
407, with recovery groups	125	73	34	107
408, without recovery groups	381	88	34	122
408, with recovery groups	205	112	34	146
410 <sup>d</sup>	19	64	34	98
411 <sup>d</sup>	9	83	34	117
412, without recovery groups	53	54	36	90
412, with recovery groups <sup>e</sup>	24	80	36	116
413, without recovery groups	52	103	36	139
413, with recovery groups <sup>e</sup>	37	127	36	163

<sup>a</sup> OECD TG 443 is based on many cohorts to assess several endpoints. The number of cohorts affects the number of animals/sex/dose. One of them requires the mating of the F1 generation to produce the F2 generation. <sup>b</sup> For OECD TG 416 tests, the average calculation excludes a study for which one dose group and control was added to an EPA-required study in order to fulfill a REACH requirement. Only the one test group for REACH and its control group are counted in our analysis. Because this is not representative of the full OECD TG 416 study, it is excluded from the average calculation. The resulting calculation is 111,359 animals/43 tests = 2,590 avg. animals/test. <sup>c</sup> OECD TG 421 and 422 had two methods of dose range-finding tests: one with mated animals and offspring, and one with unmated animals and no offspring. In these paired numbers, the first number is when the dose range-finding test included offspring and the second number is when it did not. <sup>d</sup> Recovery groups are not broken out separately for dermal (OECD TG 410 and 411), because only four OECD TG 410 tests and no OECD TG 411 tests included recovery groups. <sup>e</sup> Average shown for inhalation exposure is for two recovery groups. Five OECD TG 412 and 12 OECD TG 413 tests used four recovery groups. Recovery groups typically used either 5 or 10/sex/group, so the average for four groups would be about 20 or 40 more animals, respectively. Three tests used timed recovery groups, which added 120 to 150 animals. Seven OECD TG 412 and five OECD TG 413 included a lung burden/bronchoalveolar lavage group, adding 12-24 more animals for the OECD TG 412 and 24 to 60 for the OECD TG 413.

inhalation studies. These groups may be added at the registrant's discretion or be required by ECHA in a compliance check or test proposal decision.

**Recovery groups:** Recovery groups are animals held for a specified period after treatment ends to assess recovery of the animals, persistence of effects, and delayed occurrence of toxicity. If oral and dermal recovery groups are used, the OECD test guidelines recommend two groups: a control group and a high-dose group. For inhalation recovery groups, the recommendation is four recovery groups, one for each dose (OECD, 2018a,b).

Of the studies counted here, about 35% of sub-chronic tests and 30% of short-term tests included recovery groups, and this

was consistent across oral, dermal, and inhalation studies. Two groups were typical: a control recovery group and a high-dose recovery group. Each group usually had 5 animals per sex or 10 animals per sex, with 5 animals per sex more common. This translated to an additional 20 or 40 animals for studies that used recovery groups. About 25% of inhalation recovery groups had four groups, adding 40 or 80 animals.

**Bronchoalveolar lavage and lung burden groups:** Before the 2018 revision of OECD TG 412 and TG 413, special groups to assess lung burden and bronchoalveolar lavage fluid could be added to inhalation repeated-dose tests if exposure to particle aerosols and nanomaterials was a concern. The 2018 revision

**Tab. 6: Average animals/sex/dose compared with guideline recommendation**

OECD test guideline	Guideline recommendation for number of males and females per dose group	Avg. no. M/F per dose	Most frequently used no. M/F per dose
<b>Reproductive toxicity</b>			
415	Sufficient number of mating pairs to yield at least 20 pregnant females	23/24	24/24
416 F0	Sufficient number of mating pairs to yield at least 20 pregnant females	26/26	Tie: 24/24, 25/25, 28/28
416 F1	20 pairs	26/26	24/24
443 F0	Sufficient number of mating pairs to yield at least 20 pregnant females.	25/25	25/25
443 F1	20 pairs	21/21	20/20
421	At least 10 males and 10 females/dose group	11/11	10/10
422	At least 10 males and 10 females/dose group <sup>a</sup>	11/11	10/10
<b>Developmental toxicity</b>			
414-rat	Sufficient number of females to result in approximately 20 female animals with implantation sites at necropsy	24F	24F
414-rabbit	Sufficient number of females to result in approximately 20 female animals with implantation sites at necropsy	23F	22F
<b>Repeated-dose toxicity</b>			
408	At least 10 females and 10 males at each dose level	10/10	10/10
411	At least 10 females and 10 males at each dose level <sup>a</sup>	10/10	10/10
413	At least 10 males and 10 females per dose group <sup>a</sup>	10/10	10/10
407	At least 5 females and 5 males at each dose level <sup>a</sup>	5/5	5/5
410	At least 5 females and 5 males at each dose level <sup>a</sup>	5/5 or 10/10 (bimodal)	10 used 5/5; 8 used 10/10
412	At least 5 males and 5 females per dose group <sup>a</sup>	6/6	5/5

<sup>a</sup> Also recommends considering recovery groups of 5/sex/dose for control and high dose. For OECD 412 and 413, recommendation is for recovery groups for all dose groups, and additional groups for lung burden analysis if substance is a poorly soluble aerosol. The numbers here do not include these extra groups.

no longer allows special groups for the bronchoalveolar lavage analysis. Instead, it requires it for all tests but not on special groups; the analysis is done as part of the routine analyses on main test animals. Lung burden analysis, allowed if lung retention is a concern, is still done with special groups. Lung burden groups add 20 to 60 animals per test, depending on the number of time intervals used.

Of the studies in our animal count, seven short-term inhalation studies (OECD TG 412) and five sub-chronic inhalation studies (OECD TG 413), all conducted in 2018 or earlier, included bronchoalveolar lavage groups. Two sub-chronic studies included lung burden groups. Our review of ECHA's compliance check and test proposal decisions indicates lung burden groups may become more common. In the 2018-2022 decisions, ECHA requested these extra groups in 20 decisions.

### 3.2 Average number of animals for tests

Table 5 summarizes the average number of animals per test, both with and without a dose range-finding test. Both cases are shown because not all studies used a dose range-finding test. The aver-

age for the main test includes the following animals:

- Main test animals, i.e., the animals used for the main guideline study.
- All generations in multi-generation tests (OECD TG 414, 415, 416, 443, 421, 422).
- Animals in extra groups used to evaluate effects beyond the basic guideline study. These primarily occurred for repeated-dose tests, some of which included recovery groups, lung burden groups, and bronchoalveolar lavage groups.
- Spare animals.

The administration protocols for OECD TG 421 and OECD TG 422 are similar, which is why they have essentially the same average.

Table 6 compares the average number of animals per dose with the number of animals recommended in the OECD test guideline. For reproductive toxicity and developmental toxicity tests, the guideline is generally “a sufficient number to obtain 20 pregnant females”, and the data show that, in practice, this is 25 pairs for reproductive toxicity tests (OECD TG 443) and 22-24 females for developmental toxicity tests (OECD TG 414).

**Tab. 7: Animal estimate for pending reproductive toxicity, developmental toxicity, and sub-chronic toxicity tests**

These are new tests required by an authorized testing proposal or a compliance check, but they were not yet present in the dossiers at the time of our analysis. Method details for determining pending tests are in Section 2.4. Avg. animals/test is from Table 5. The average includes a dose range-finding (DRF) test, because most studies have included DRF tests as the dose selection method (Section 3.1.2).

OECD test guideline	No. of authorized proposals <sup>a</sup>	No. of compliance check decisions requiring test <sup>a</sup>	Avg. animals/test, incl. DRF	Estimated total animals
443 without F2 extension	54	82	1,826	248,336
443 with F2 extension	18	25	2,734	117,562
421/422	not required	211	604	127,444
414-rat	87	269	1,460	519,760
414-rabbit	51	147	1,095	216,810
408/413	77	260	122 <sup>b</sup>	41,114
<b>Total tests pending</b>	<b>287</b>	<b>994</b>		<b>1,271,026</b>

<sup>a</sup> Numbers obtained from <https://www.echa.europa.eu/information-on-chemicals/dossier-evaluation-status>, downloaded 29 January 2023.

Numbers exclude tests annulled or withdrawn by ECHA on appeal, tests related to registrations that are no longer valid, and tests no longer necessary because ECHA accepted an adaptation instead of the test. Read-across decisions were also excluded unless the test was required on the registered substance as part of a category read-across decision. <sup>b</sup> Sub-chronic test requests are about 90% OECD TG 408 (oral exposure) and 10% OECD TG 413 (inhalation exposure). For simplicity, average animals/test is set to that for OECD TG 408, without recovery groups. For reference, the average animals/test for OECD TG 413 without recovery groups and including DRF is 139 (Tab. 5).

### 3.3 Pending animal tests

New reproductive toxicity, developmental toxicity, and sub-chronic toxicity tests are under way or soon to be conducted as a result of ECHA compliance checks and ECHA authorizations of registrant testing proposals for these test types. These pending tests were estimated for ECHA decisions from 2018–2022 (Section 2.4), the period likely to have outstanding tests. Table 7 shows the results.

### 3.4 Uncertainty analysis

Uncertainties relate to assumptions incorporated into the algorithm for extracting the studies; the estimation techniques for unknown data, such as number of pups; and the quality of the data itself. The following known factors may have led to underestimates or overestimates:

*Restriction of studies to reliability 1:* Our algorithm excludes studies with reliability of 2. A manual review of 100 reliability 2 studies for oral repeated dose toxicity identified two studies that were for REACH but were reliability 2 because of study problems, and three more that potentially were for REACH but with a reference type (“other company data”) that made it uncertain. Assuming all five were for REACH, that suggests a potential undercount of 5% of REACH studies that were reliability 2. eChemPortal had 2,726 reliability 2 studies for the endpoints we counted. Assuming the 5% applies across all endpoints, we potentially missed 136 reliability 2 studies, which is about 3% of the total studies counted. Potential impact: Underestimate of about 90,000 animals.

*Exclusion of studies with report date < 2010 or study year < 2009:* Rerunning the algorithms to allow 2008 studies and allow the report year to be 2008 or 2009 yields an additional 61 screening studies (OECD TG 421/422), 9 reproductive toxicity

studies (OECD TG 415/416), 19 developmental toxicity studies (OECD TG 414), and 43 repeated-dose toxicity studies (OECD TG 407/408). Potential impact: Underestimate of about 80,000 animals.

*Estimate of spare animals as 10% of main test animals (parent animals for multigeneration tests):* The few dossiers with this data typically reported 10% to 25% spare animals. The 10% estimate is the conservative end of that range, but the data were so few that it may not be representative. We do not know if 10% is an over- or underestimate. At 10%, total number of spares for existing studies is 39,089, or 1.3% of the total animal count. Potential impact: Unknown, but likely minimal.

*Pup estimation method:* The key parameter for estimating number of pups is average number of pups per litter, based on litter sizes reported in the dossiers, rounded to the nearest whole number. The data underlying the averages are good, and we are confident in the averages. The issue is the rounding error. For rabbits, the average is 9.0, so there is no rounding error. For rats, however, the average is 11.6, which is rounded to 12. That rounding up is equivalent to 71,194 pups. Potential impact: Overestimate of about 70,000 animals.

The other parameter in the pup estimate is average fertility index (number of paired females who become pregnant). The average calculated from dossier data is 92% for rats and 94% for rabbits. It is adjusted down in both cases to 90% to account for gestation deaths. This adjustment introduces a potential error in the other direction. If actual fertility indices were used, the number of pups would increase by 27,881. Likely, the percentage is somewhere between 90% and the actual average. Potential impact: Underestimate of 15,000–24,000 animals.

*Inclusion of too many OECD TG 415 and TG 416 studies:* Our count includes 44 OECD 416 studies and 13 OECD TG 415 stud-

ies. Comparison of these studies with compliance check and test proposal decisions shows few matches. No OECD TG 416 tests were authorized by ECHA after 2012. Instead, from 2012–2015, OECD TG 416 decisions were deferred to the European Commission, pending a decision on whether to continue with OECD TG 416 or to require the new OECD TG 443 method instead (ECHA, 2013, 2014). REACH was amended in March 2015 to specify OECD TG 443 as the required method, and OECD TG 416 tests were not accepted thereafter unless they started before March 2015. The OECD TG 416 studies in our count may be REACH studies that were done without advance authorization, an issue that was investigated by ECHA for many test types performed without proposals from 2009 to about 2013 (ECHA, 2015). ECHA reported that most registrants they queried about these tests responded that the test had a non-REACH purpose. Had registrants answered that the test was for REACH, however, they would have opened themselves to legal action, so it is unclear how to interpret the responses. Several OECD TG 416 studies in our count were identified in the ECHA (2015) report as tests whose test proposal evaluations were terminated because the test was already ongoing.

ECHA's third report on alternative methods, which covers through July 2016, includes a count of new experimental tests conducted since 2009 (ECHA, 2017c). At that time, ECHA found 23 OECD TG 416 studies. Since OECD TG 416 studies could no longer be started after March 2015, this suggests that 23 may be the maximum number of OECD TG 416 tests potentially done for REACH. Possibly, the 21 additional tests we found were existing studies entered into dossiers after ECHA completed their third report. Potential impact: Overestimate of about 60,000 animals.

OECD TG 415 has never been a standard REACH requirement, but registrants did submit test proposals for OECD TG 415, so it was a test considered at the time. Eleven of the 13 OECD TG 415 tests were done between 2010 and 2013, suggesting these were for REACH. The total animal count for all 13 tests is only 18,846, so any impact of overcounting a few is slight. Potential impact: negligible.

*Inclusion of non-REACH and duplicate studies:* This is difficult to estimate. To detect non-REACH studies, our algorithm flagged tests that did not reference an OECD or EU test guideline and checked for appropriate species. It also extracted for manual review several dossier fields where registrants sometimes note if a test is performed for a purpose other than REACH. However, there is no requirement for registrants to specify this. To detect duplicate studies, the algorithm compared test material with the registered substance and flagged any mismatches or blank entries for review. It also extracted the study period and in-life dates, which often contained full dates (day, month, year), details that allowed tests to be sorted by those fields, with matches almost always indicating duplicate studies. We believe we caught most duplicates, but duplicates may have been missed for studies where the test material was not identified (such tests were included if they met narrow criteria indicating they were likely for REACH – see Section 2.2) and only a year was given. In those cases, data were not detailed enough to allow study comparisons. When we compare our saved sub-chronic, developmental, and full repro-

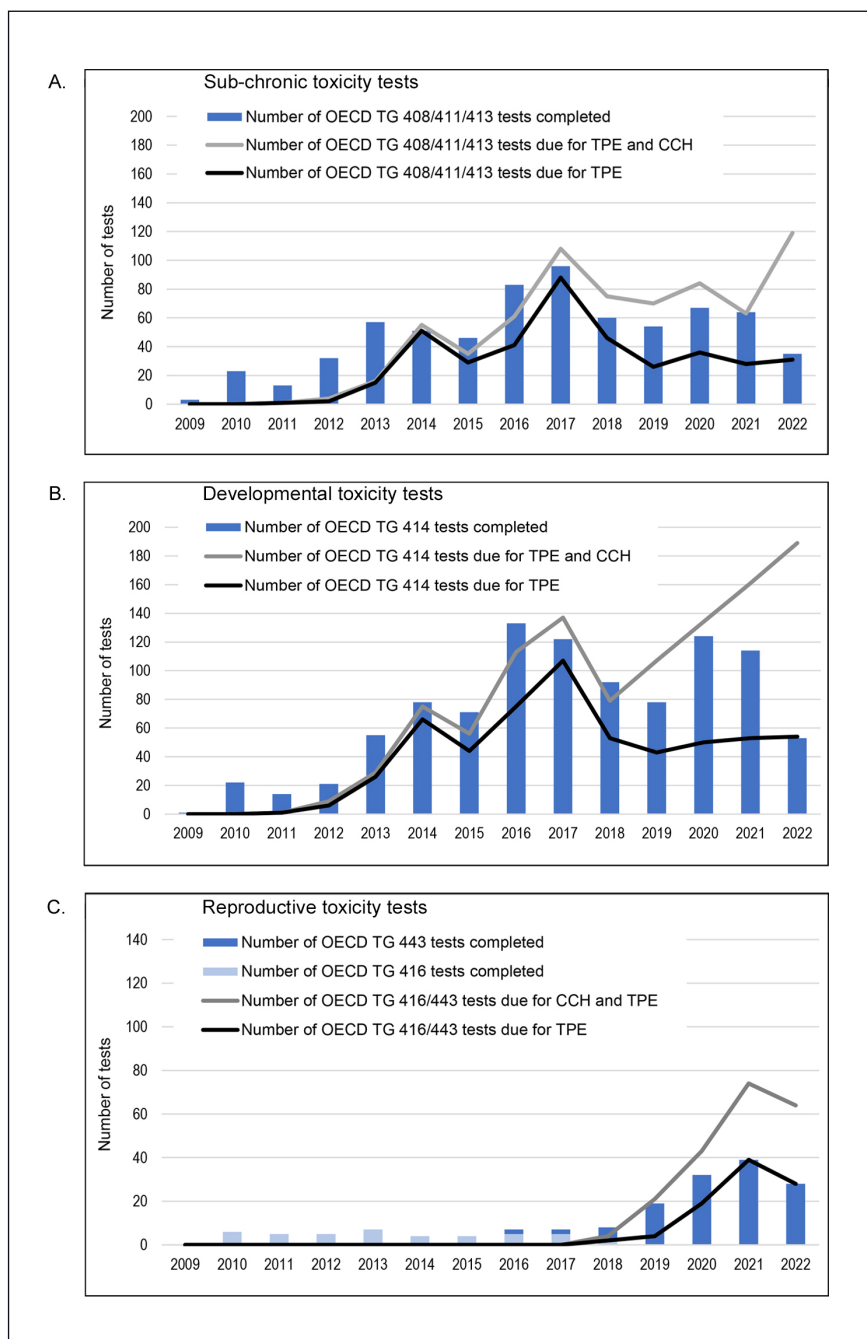
ductive toxicity studies, all of which require test proposal authorizations, against the ECHA list of test proposal evaluation decisions and compliance check decisions, we find no matching decision for 3% to 18% of our saved studies, depending on endpoint (excluding OECD TG 416 and TG 415, which are a special case addressed above in this analysis). The miss rate drops to 3% to 7% if we exclude the years 2009–2013, when test proposals were not always submitted. These unmatched studies are either non-REACH studies, duplicates, or tests done without a proposal. Figures 1A and 1B show peaks of unmatched studies in 2010 and 2013, the REACH deadlines for registering Annex IX and X substances, suggesting that missing test proposals explain most unmatched studies in this period. That leaves 3% to 7% unmatched studies for 2014–2022. If we make the conservative assumption that all are non-REACH or duplicate tests, the total number of animals for the unmatched tests for endpoints requiring authorization by ECHA is 95,280. If we then assume that OECD TG 407, 421, and 422 tests, which require no authorization, have a similar rate of duplicate and non-REACH tests (assume 3%–7%), this adds about 34,000 to 77,000 animals. Potential impact: Overestimate of 130,000 to 170,000 animals.

Comparing the tests included in our count with the tests required by ECHA test proposal evaluations and compliance check decisions provides insight into the scale of over- or underestimates. In theory, the number of tests we found for sub-chronic toxicity, developmental toxicity, and full reproductive toxicity endpoints should be close to the number of tests required by ECHA test proposal evaluations and compliance checks for those endpoints. Figure 1 shows this comparison from 2009–2022. For each year, it shows the number of ECHA test requests due each year and the number of tests in our animal count with that study year. Note that the figure refers to authorized proposals as TPE (test proposal evaluation decision) and compliance check decisions as CCH (compliance check decision), in keeping with ECHA terminology. The TPE and CCH test requests in the figures exclude requests that were fulfilled by an acceptable adaptation instead of a test and requests that are no longer applicable. A test may no longer be applicable because the request was withdrawn by ECHA or annulled by the Board of Appeals, the registration is no longer valid, or the registrant has ceased manufacture of the substance or downgraded the tonnage.

In Figures 1A and 1B for sub-chronic toxicity tests and developmental toxicity tests, respectively, the charts for each endpoint can be roughly divided into three periods:

- 2009–2013, when we have more tests than TPE/CCH test requests, an overcount that may be explained by tests performed without testing proposals. Test peaks are visible at the REACH 2010 and 2013 deadlines.
- 2014–2018, when the match is good, and minor divergences between number of tests due each year and number found for each year are explained by tests being completed earlier or later than the deadline year, so the actual test appears in a year before or after the CCH or TPE due date.
- 2019–2022, when we have fewer tests than TPE/CCH test requests, either because we discovered the tests after finalizing our count, the tests are ongoing but not yet present in the dos-





**Fig. 1: Number of tests completed each year compared with number of tests due each year from ECHA test proposal (TPE) and compliance check (CCH) decisions**

The lines show the number of tests due each year, required by ECHA test proposal authorizations (TPEs) or compliance checks (CCHs), excluding test requests fulfilled by an adaptation or no longer applicable. The upper line shows the number due for both TPEs and CCHs combined; the lower line shows number of tests due just for TPEs, to allow comparison between TPE and CCH test requests. The bars show the actual number of studies completed that year. Ideally, the bars would reach the upper line, but studies were often completed a year or two earlier or later than the due date. The figure also shows a low submission of test proposals from 2009–2013. For reproductive toxicity (Fig. 1C), tests were further delayed until 2017, pending determination of the preferred test method (OECD TG 416 or 443). The large gap between the upper line and bars from 2019 forward reflects CCH and TPE tests that are not in our animal count, either because we discovered them after finalizing our count or the test was not present in the dossier. Study years are the years reported for the studies in the dossiers in the public ECHA database. Due dates for the tests are from the ECHA Dossier Evaluation Status list.

sier, or the dossiers have no test and have “Follow-up” status, ECHA has not yet accepted the dossiers as compliant. There can be a time lag of several years for TPE and CCH tests to appear in the dossiers. Most of these missing tests are included in the animal count as pending tests (Section 3.3).

Figure 1C, for reproductive toxicity, shows little activity before 2018, reflecting both the lack of testing proposals in the early years and the suspension of all test authorizations for OECD TG 416 in 2012, until the European Commission began processing deferred decisions in 2017, generally issuing requests for OECD

TG 443 (ECHA, 2018). After 2018, OECD TG 443 tests started, and the match between TPE requests and our OECD TG 443 test count is good. The gap with CCH test requests has the same explanation as that for the other endpoints.

The estimate of pending tests also has uncertainty, related to test requests that may be fulfilled by adaptations. Potential adaptations cannot be estimated for any given future year, because historical data indicate the percentage of requests fulfilled by adaptations varies widely from year to year. The overestimate is probably at least 10% based on historical data.

**Tab. 8: Comparison of the number of expected registrations according to Pedersen et al. (2003) and Rovida and Hartung (2009) with the ECHA data on current number of registered substances<sup>a</sup>**

Pedersen et al. (2003) and Rovida and Hartung (2009) included intermediates imported or manufactured in a quantity  $\geq 1000$  tons/year.

	> 1 t/y	> 10 t/y	> 100 t/y	> 1000 t/y	TOTAL
Pedersen et al. (2003) estimate	19,200	4,977	2,461	2,704	29,342
Rovida and Hartung (2009) estimate, considering market increase	44,632	11,570	5,721	6,286	68,209
Registered substances (Nov. 2022) <sup>a</sup>	8,695	5,522	3,834	2,355	20,406

<sup>a</sup> <https://echa.europa.eu/registration-statistics>

Many of these uncertainties could be reduced significantly by including new dossier fields that indicate if the test was done for REACH; if the test was done for the registered substance; and total number of pups, including those culled.

## 4 Discussion

When REACH was proposed in the early 2000s, Pedersen et al. (2003) assessed the additional testing needs that would result from REACH in a report prepared for the European Commission. That report necessarily involved assumptions of expected registrations, amount of existing testing data, use of the 3Rs (replace, reduce, refine) principle for minimizing animal use, and acceptance of those methods by registrants and regulators. Based on the Pedersen et al. (2003) assumptions, van der Jagt et al. (2004) estimated that the full implementation of REACH would sacrifice 2.6 million animals, while Rovida and Hartung estimated 54.4 million animals considering the market increase. The present report measured 2.9 million animals already present in the ECHA database plus 1.3 million animals in ongoing studies, for a total of 4.2 million animals. This is for only three categories of human health endpoints – repeated-dose toxicity, reproductive toxicity and developmental toxicity – and excludes the animals that will be added after the conclusion of the compliance checks and the extension of the technical completeness checks to all existing submissions.

### 4.1 Number of registered substances

The ECHA numbers of registered substances in the different tonnage bands are taken from the REACH registration statistics that are regularly published in the ECHA database in the section about “Information on chemicals” (update of 30/11/2022<sup>9</sup>).

A major difference between the real situation and past estimates is the total number of registered substances, which is much lower than estimated in the Pedersen et al. (2003) paper or Rovida and Hartung (2009) (Tab. 8). Comparing the total number of registered substances, the actual number represents 69.5% of the number considered by Pedersen et al. (2003) and 30% of the number considered by Rovida and Hartung (2009).

The reasons for the large difference between the predicted number of substances and number of registered substances are unknown, but the economic crisis of 2008 and the registration cost may have been important factors. Companies may have preferred to quit the manufacturing or the importation of a substance to avoid the registration, in particular when similar chemicals were available as alternatives. Cost considerations may have also caused UVCB substances (unknown or variable composition, complex reaction products or biological substances) to be grouped together under the same registration, by broadening the definition of the boundary composition.

Going forward, the number of registrations will change as new substances are registered and others are retired. New substances may require new animal testing.

### 4.2 Number of animals per test

Table 9 compares average animals/test used in Rovida and Hartung (2009) and in van der Jagt et al. (2004) with the averages calculated from REACH tests from 2009–2022. The van der Jagt et al. (2004) estimation was reached mainly by consulting test laboratories that were performing the tests. These numbers did not include offspring, because at that time it was not requested. Rovida and Hartung (2009) based the estimation on the corresponding OECD test guideline and included offspring. Neither paper considered the dose range-finding studies. The last column contains the average animals/test calculated for the tests reported in the REACH registration dossiers and includes dose range-finding studies, which are part of the study in most of the cases (Section 3.1.2). Van de Jagt et al. (2004) assumed that few dose range-finding tests would be needed because historical data would be available to select doses. In practice, dose range-finding tests were often used even when historical data were available, in combination with them. Note that even our calculation underestimates dose range-finding animals, because our count does not include a preliminary range-finding test for the main dose range-finding test. This is sometimes performed on fewer animals or for a shorter period of time in order to define the best doses, with the highest causing a little toxicity to the animals and the lowest being completely safe. This is important to avoid all the animals dying before the study is concluded.

<sup>9</sup> [https://echa.europa.eu/documents/10162/2741157/registration\\_statistics\\_en.pdf/58c2d7bd-2173-4cb9-eb3b-a6bc14a6754b?t=1649160655122](https://echa.europa.eu/documents/10162/2741157/registration_statistics_en.pdf/58c2d7bd-2173-4cb9-eb3b-a6bc14a6754b?t=1649160655122). See “Breakdown by Tonnage Band” in the “Detailed Registration Statistics” section.



**Tab. 9: Comparison of average number of animals used in van der Jagt et al. (2004) and Rovida and Hartung (2009) and average number calculated from REACH tests from 2009-2022 and reported in Table 5**

Test guideline	Avg. number of animals/test in van der Jagt et al. (2004)	Avg. number of animals/test in Rovida and Hartung (2009)	Avg. number of animals from Tab. 5, excl. DRF animals	Avg. number of animals from Tab. 5, incl. DRF animals
OECD TG 407	50	40	44 <sup>a</sup>	78 <sup>a</sup>
OECD TG 410	50	40	64 <sup>a</sup>	98 <sup>a</sup>
OECD TG 412	50	40	54 <sup>a</sup>	90 <sup>a</sup>
OECD TG 408	80	80	88 <sup>a</sup>	122 <sup>a</sup>
OECD TG 411	NR <sup>b</sup>	80	83 <sup>a</sup>	117 <sup>a</sup>
OECD TG 413	NR <sup>b</sup>	80	103 <sup>a</sup>	139 <sup>a</sup>
OECD TG 421	80	560	566	802/598 <sup>c</sup>
OECD TG 422	–	412	572	808/604 <sup>c</sup>
OECD TG 414 (rat)	100	784	1,161	1,460
OECD TG 414 (rabbit)	100	560	880	1,095
OECD TG 416	448	3,200	2,590	3,098
OECD TG 443	–	–	1,318/2,226 <sup>d</sup>	1,826/2,734 <sup>d</sup>

<sup>a</sup> Excludes recovery groups. About 30% of short-term repeated-dose tests and 35% of sub-chronic tests included recovery groups, which typically added a total of 20 or 40 animals to the test. <sup>b</sup> Not reported, due to the variability of interim animals, i.e., animals that are added and sacrificed before the end of the exposure period. <sup>c</sup> In these paired numbers, the first number is when the dose range-finding test used mated animals (with offspring) and the second number is when it used unmated animals (no offspring). <sup>d</sup> The first number is OECD TG 443 without the F2 generation; the second number is with the F2 generation.

DRF = dose range-finding

Recovery groups and other extra groups included in studies also were not considered in the earlier predictions as they are not part of the main study, but they are animals sacrificed for the purpose of the experiment and must be counted according to Directive 2010/63.

Another difference between the real numbers and the estimation is that Rovida and Hartung (2009) considered the possibility to perform a limit test at 1,000 mg/kg for substances that are not expected to be toxic. A limit test is a study performed only at the highest dose, reducing the number of animals that are necessary for a regular test based on four doses. Even though the OECD test guidelines allow a limit test, this is rarely used because one cannot be sure that this high dose will cause no effect, and it blocks the possibility to derive a No Observed Effect Level (NOEL), which is necessary for the CSA.

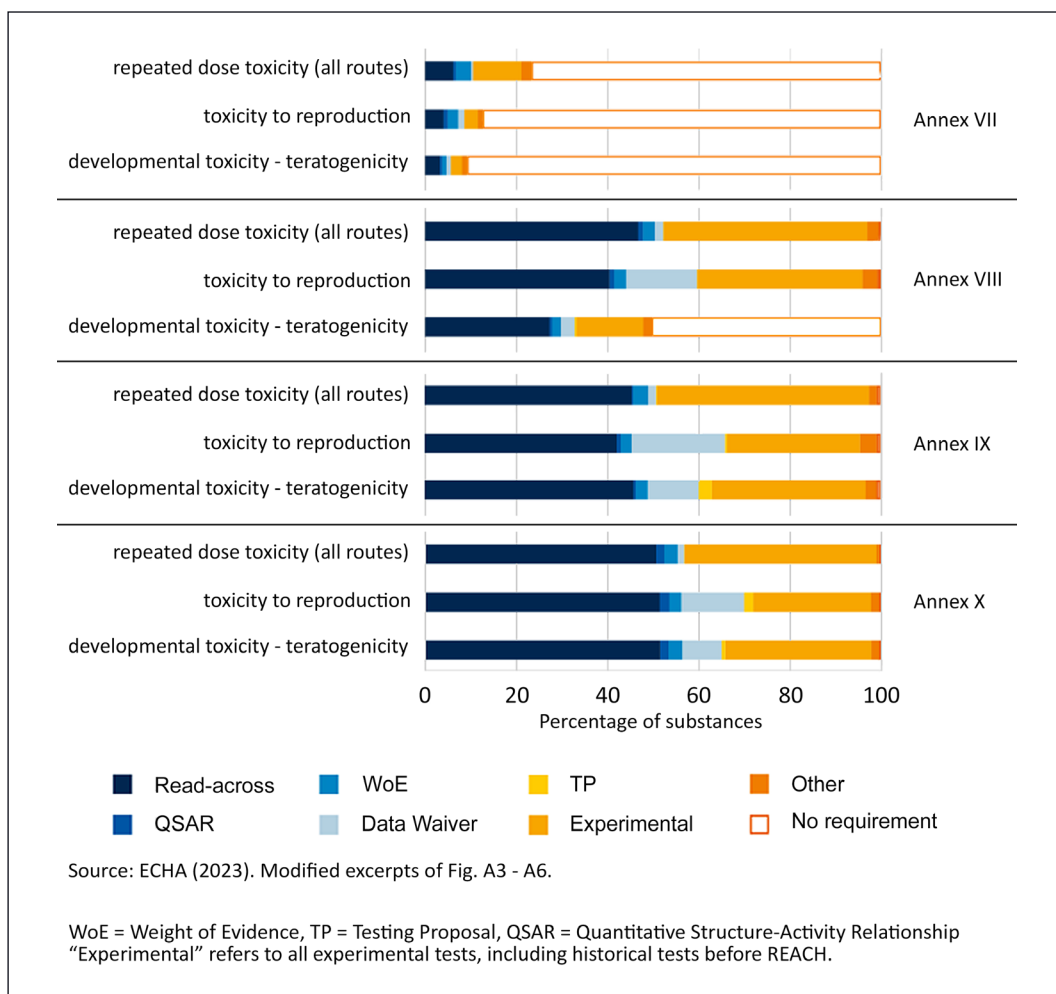
#### 4.3 Use of 3Rs

Figure 2 is extracted from a figure in the latest ECHA report on the use of alternatives to testing on animals for REACH (ECHA, 2023). Compared to the figure in the ECHA report, it shows only the data regarding reproductive toxicity, developmental toxicity, and repeated-dose toxicity in the registrations under Annexes VII, VIII, IX and X. Disregarding Annex VII, for which these tests are performed only in exceptional circumstances, this graph reveals that read-across is applied by the registrants at a rate of approximately 50%. Real numbers are not shown in the

ECHA figure, but apparently these data are in line with what was predicted by Pedersen et al. (2003).

The proportions of Figure 2 will change at the end of the compliance checks. For the endpoints that are considered in the present assessment – repeated-dose toxicity, reproductive toxicity, and developmental toxicity – no accepted QSAR method can cover the endpoint as is requested in the REACH Regulation (ECHA, 2017c). QSARs have proven to be effective in support of the read-across approach or as an additional assessment when applying the weight-of-evidence approach (Rovida et al., 2020).

The read-across rejection rate during compliance checks has been very high, often due to an unsatisfactory justification (ECHA, 2016). In the evaluation of read-across studies, ECHA applies the conditions described in the Read-Across Assessment Framework (RAAF) (ECHA, 2017d), which includes toxicokinetic similarity as a requirement. For UVCB substances, ECHA asks for toxicokinetic information for all components, and only minor differences, which are intrinsic to the definition of a UVCB substance, stop the applicability of read-across. In the 2021 report on the operation of REACH and CLP (ECHA, 2021), ECHA reports that only 25% of the read-across studies evaluated during compliance checks have been accepted, triggering the request for the standard information requirement, i.e., a new animal test, to cover the endpoint. Registrants may still provide a read-across study, and ECHA generally provides guidance about



**Fig. 2: Data retrieved from the latest ECHA report on the use of alternatives to testing on animals for the REACH Regulation (ECHA, 2023)**

The experimental data include both existing historical tests and tests performed for REACH.

what would be required to make the read-across acceptable, but comparison of our studies with compliance check test requests indicates that most registrants provide the requested animal test. In our review of 429 compliance check decisions of 2018-2019 requesting new sub-chronic, developmental, or reproductive toxicity tests and for which the deadline for submitting the test had passed, registrants provided a new test for 67% and an adaptation (usually read-across) for 26%; the other 7% were unclear. Providing an adaptation instead of the test risks a second rejection that would include the payment of a fine in addition to the confirmation that the new test must be performed.

Our analysis focuses on the number of animals used in new tests by counting them in the ECHA database. The assessment of the proportion of alternatives used by registrants to prepare dossiers is out of our scope. Further details on the application of 3Rs in the REACH registration dossiers are presented by Taylor (2018).

#### 4.4 The toll of REACH on animals continues

Many clues indicate that the toll of REACH on animals is likely to increase.

To the estimate of 2,902,807 animals present in the ECHA database (Tab. 2), we added the estimate of 1,271,026 animals that are in use in tests already approved by ECHA but not yet recorded in the dossiers (Tab. 7), arriving at an estimated total of 4,173,833 animals already killed or being killed for REACH purposes. This is for only three categories of human health endpoints – for repeated-dose toxicity, reproductive toxicity, and developmental toxicity – and excludes all other human health endpoints and all ecotoxicity endpoints. The tests for the other categories related to human health – in practice, acute tests and genotoxicity tests – use fewer animals, but ecotoxicity tests require hundreds of fish per test. Fish are vertebrate and sentient animals. Application of other tests such as chronic and carcinogenicity tests for human toxicity are excluded from our assessment and they should be considered in future, while long-term or reproductive toxicity to birds was not found in the dossiers and can be disregarded.

The preferred option recognized for the adaptation of standard information requirements and waiving of animal tests is the use of read-across, but we have already mentioned that ECHA reports only 25% of read-across approaches have been accepted during compliance checks (ECHA, 2021). If all dossiers undergo the



compliance check procedure, we can expect a steep increase in the request for new animal tests. In some cases, new animal tests are also requested when available experimental studies are rejected because they were performed with an old OECD test guideline that is missing some information requested today or because the analytical characterization of the test item was not sufficiently detailed (personal experience; no publication is available).

The situation is even more concerning because continuous updates to the legislation may make old dossiers not compliant. Regulation EU 2022/477 amending Annexes VI to X of the REACH regulation and in force since 14 October 2022 explicitly limits the possibility to waive reproductive toxicity tests and developmental toxicity tests on the second species in substances registered in the tonnage band  $\geq 1,000$  tons/year. This Regulation also expands the request for animal tests for the assessment of genotoxicity properties and aquatic toxicity for substances registered at the lowest tonnage band. This is out of the scope of our analysis, which is focused only on other endpoints.

A new important revision of REACH is under discussion, and the draft is expected by the end of 2023. Among many other issues, this update will consider the registration of polymers that may present toxicity and the expansion of the CSA to include substances registered in Annex VII. The ramifications of the proposed revisions are discussed in detail in Rovida et al. (2023), also in this issue. If the general approach of REACH is not changed through a serious opening to the acceptance of new approach methodologies (NAMs) and other advanced methods that can predict the toxicity of substances with high confidence and without using animals, the toll of REACH on animals is set to increase greatly.

## 5 Conclusions

This analysis presents the number of animals that have been used for compliance with the REACH regulation for three test categories: repeated-dose toxicity, developmental toxicity, and reproductive toxicity. A direct count of animals from the REACH dossiers shows about 2.9 million animals have been used and an additional 1.3 million animals are being used in ongoing studies, for a total of 4.2 million animals. The total for these three categories alone exceeds the expectation of the European Commission in 2004, during the approval phase of the REACH regulation, of 2.6 million animals as the total animal toll for the implementation of the new regulation.

The difference is primarily because the 2004 forecast excluded pups, which was an accepted counting method at the time. However, pups make up the majority of animals sacrificed for REACH. Of the 2.9 million animals used to date for systemic toxicity tests for human health, about 2.1 million were pups. Secondly, the original forecast assumed registrants would use existing data for selecting suitable doses rather than conduct dose range-finding tests on animals. Of the study reports that included the dose selection method, however, most reported use of a dose

range-finding test either alone or together with existing data. These tests have used about 350,000 animals to date, not including those related to the ongoing studies.

Reviewing use of alternative methods to reduce animal numbers was beyond the scope of our analysis, but our analysis appears to confirm problems with the read-across approach under REACH. Read-across guidance has been available since 2017 (ECHA, 2017b), but ECHA reports that only 25% of the read-across approaches have been considered valid in compliance checks (ECHA, 2021), triggering ECHA requests for the standard information requirement, often animal tests. Registrants can still respond with a non-animal adaptation, but comparison of our studies with compliance check test requests indicates that most registrants provided the requested animal test. Given that read-across is the only alternative method available for the most animal-intensive tests, its failing use requires further action to resolve the underlying issues.

According to REACH Article 1, a purpose of the REACH regulation is the promotion of alternative methods for assessment of hazards of substances, while the request for new animal tests should be considered only as a last resort. From this perspective, our assessment has a double use: It should serve to increase awareness of the high number of animals that each systemic toxicity test needs and support ECHA in its efforts to require new tests only when there is a clear benefit for the improvement of human and environment health. It also demonstrates the possibility to monitor animal use, which is important for measuring the impact of REACH in this area, and efforts by ECHA and the European Commission to increase use of alternative strategies. It is clear from this paper and ECHA (2020) that counting REACH animals is currently a laborious project, even with algorithms to help. A recommendation is to add or modify fields within the ECHA database to allow an effective count that ECHA could periodically update.

The European Commission periodically updates the REACH impact assessment in terms of benefits for human health and the environment or the consequences on the economy<sup>10</sup>. A new important revision of REACH is expected soon, as requested in Article 138 of REACH, which states that the “*Commission may present legislative proposals to modify the information requirements for substances manufactured or imported in quantities of one tonne or more up to 10 tonnes per year per manufacturer or importer, taking into account the latest developments, for example in relation to alternative testing and (quantitative) structure-activity relationships ((Q)SARs).*” Our hope is for an expansion in the use of NAMs and other alternative strategies for the preparation of the REACH registration dossiers and in general for a reliable hazard and risk assessment of chemicals.

## References

EC – European Commission (2001). White Paper – Strategy for a future Chemicals Policy. COM(2001) 88 final.

<sup>10</sup> [https://ec.europa.eu/environment/chemicals/reach/background/i\\_a\\_en.htm](https://ec.europa.eu/environment/chemicals/reach/background/i_a_en.htm)



- ECHA (2013). Evaluation under REACH: Progress report 2012. ECHA-13-A-01-EN. [https://echa.europa.eu/documents/10162/17221/evaluation\\_report\\_2012\\_en.pdf/fa360388-4c23-4816-90be-09812061e12f](https://echa.europa.eu/documents/10162/17221/evaluation_report_2012_en.pdf/fa360388-4c23-4816-90be-09812061e12f)
- ECHA (2014). Evaluation under REACH: Progress report 2013. ECHA-14-A-01-EN. doi:10.2823/22707
- ECHA (2015). Survey results – Analysis of higher tier studies submitted without testing proposals. ECHA-15-R-13-EN. doi:10.2823/11412
- ECHA (2016). Report on the Operation of REACH and CLP 2016. ECHA-16-R-08-EN
- ECHA (2017a). Guidance on Information Requirements and Chemical Safety Assessment – Chapter R.7a: Endpoint specific guidance. ECHA-17-G-18-EN. doi:10.2823/337352
- ECHA (2017b). Evaluation under REACH: Progress report 2016. ECHA-17-R-03-EN. doi:10.2823/588707
- ECHA (2017c). The use of alternatives to testing on animals for the REACH regulation. Third report under Article 117(3) of the REACH Regulation. June 2017. ECHA-17-R-02-EN. doi:10.2823/023078
- ECHA (2017d). Read-Across Assessment Framework (RAAF). ECHA-17-R-01-EN. doi:10.2823/619212
- ECHA (2018). Evaluation under REACH: Progress report 2017. ECHA-18-R-05-EN. doi:10.2823/76886
- ECHA (2020). The use of alternatives to testing on animals for the REACH regulation. Fourth report under Article 117(3) of the REACH Regulation. June 2020. ECHA-20-R-08-EN. doi:10.2823/092305
- ECHA (2021). Report on the Operation of REACH and CLP 2021. ECHA-21-R-06-EN. doi:10.2823/041059
- ECHA (2023). The use of alternatives to testing on animals for the REACH Regulation. Fifth report under Article 117(3) of the REACH Regulation. June 2023. ECHA-23-R-07-EN. doi:10.2823/805454
- Hofer, T., Gerner, I., Gundert-Remy, U. et al. (2004). Animal testing and alternative approaches for the human health risk assessment under the proposed new European chemicals regulation. *Arch Toxicol* 78, 549-564. doi:10.1007/s00204-004-0577-9
- OECD (2018a). Test No. 408: Repeated Dose 90-Day Oral Toxicity Study in Rodents. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264070707-en
- OECD (2018b). Test No. 412: Subacute Inhalation Toxicity: 28-Day Study. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264070783-en
- Pedersen, F., de Bruijn, J., Munn, S. et al. (2003). Assessment of additional testing needs under REACH – Effects of (Q)SARs, risk based testing and voluntary industry initiatives. Joint Research Centre Report EUR 20863.
- Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals – A report by the Transatlantic Think Tank for Toxicology (t<sup>4</sup>). *ALTEX* 26, 187-208. doi:10.14573/altex.2009.3.187
- Rovida, C., Barton-Maclaren, T., Benfenati, E. et al. (2020). Internationalisation of read-across as a validated new approach method (NAM) for regulatory toxicology. *ALTEX* 37, 579-606. doi:10.14573/altex.1912181
- Rovida, C., Busquet, F., Leist, M. and Hartung T. (2023). REACH out-numbered! The future of REACH and animal numbers. *ALTEX* 40, 367-388. doi:10.14573/altex.2307121
- Taylor, K., Stengel, W., Casalegno, C. et al. (2014). Experiences of the REACH testing proposals system to reduce animal testing. *ALTEX* 31, 107-128. doi:10.14573/altex.1311151
- Taylor, K. (2018). Ten years of REACH – An animal protection perspective. *Altern Lab Anim* 46, 347-373. doi:10.1177/026119291804600610
- van der Jagt, K., Munn, S., Tørsløv, J. et al., J. (2004). Alternative approaches can reduce the use of test animals under REACH. JRC Report EUR 21405 EN

#### Conflict of interest

The authors declare that they have no conflicts of interest.

#### Acknowledgments

This work received funding from the European Union's Horizon 2020 research and innovation program under grant agreements No. 964537 (RISK-HUNT3R) and No. 963845 (ONTOX).

#### Data availability

Worksheets containing the original data and calculations for the final count are in the supplemental .xlsx file included with this paper. Requests for data for other analyses presented in this paper may be made to Jean Knight at [jean.dora@gmail.com](mailto:jean.dora@gmail.com).